Analysis of the Effect of Reversibility Constraints on the Predictions of Genome-scale Metabolic Models

José P. Faria[1,2,3], Miguel Rocha[1], Rick L. Stevens[2,3] and Christopher S. Henry[2,3]

[1]Dep. Informatics/ CCTC, University of Minho, Portugal, [2] Computation Institute, The University of Chicago, Chicago, IL, USA, [3]Argonne National Laboratory, Argonne, IL, USA

1. Abstract

Reversibility constraints are one aspect of genome-scale metabolic models that has received significant attention recently. This study explores the impact of complete removal of reversibility constraints on the gene essentiality and growth phenotype predictions generated using three published genome-scale metabolic models: the *i*JR904, the *i*AF1260, and the *i*Bsu1103. In all three models, the accuracy in predicting essential genes declined significantly with the relaxation of reversibility constraints, while the accuracy in predicting nonessential genes increased only for the *i*JR904 and *i*AF1260 model. Additionally, the number of inactive reactions in all models declined substantially with the relaxation of the reversibility constraints. This study rapidly reveals the extent to which the reversibility constraints included in a metabolic model have been optimized, and it indicates those incorrect model predictions that may be repaired and those correct model predictions that may be broken by increasing the number of reversible reactions in a model.

2. Introduction

In recent years, Flux Balance Analysis (FBA) and genome-scale metabolic models are increasingly being used as a means of predicting the metabolic capabilities of an organism based on knowledge of the biochemical interactions taking place in the organism's metabolic pathways. These models are capable of predicting essential genes, growth phenotypes, culture conditions, and metabolic engineering strategies [1]. Additionally, the number of models available for analysis is rapidly growing. Currently, models have been published for over 20 microorganisms [2],

with new high-throughput reconstruction methods emerging capable of producing thousands of draft models in a single year [3].

One aspect of genome-scale metabolic models that has received significant attention recently is the reversibility constraints governing the direction(s) of operation for all reactions included in the model. In the first genome-scale metabolic models, these constraints were based largely on data available in biochemical databases and knowledge of pathway directionality in well-known metabolic subsystems (e.g. glycolysis) [4, 5]. More recently, methods have emerged for predicting reaction reversibility/directionality based on thermodynamics and simple heuristic rules [6-8]. Finally, methods are available for adjusting reaction reversibility/directionality constraints to fit model predictions to available experimental phenotype data [9, 10]. All of this work demonstrates the impact that small targeted changes to model reversibility constraints have on the accuracy of model predictions. Here, we explore the impact of complete removal of reversibility constraints on the gene essentiality and growth phenotype predictions generated using three published genome-scale metabolic models: the *i*JR904 [4], the *i*AF1260 [11], and the *i*Bsu1103 [6]. The *i*JR904 and *i*AF1260 are both metabolic models of *E. coli* K12. The *i*JR904 model includes 931 reactions encompassing 904 ORFs; the *i*AF1260 model is a substantial expansion over the *i*JR904, including 2059 reactions and encompassing 1260 ORFs. The *i*Bsu1103 model was created for *B. subtilis* 168 and includes 1437 reactions encompassing 1103 ORFS. The *i*Bsu1103 model was optimized using the *GrowMatch* [9] method, in contrast to the *E. coli* models, which were manually optimized. This study is part of a larger study examining the impact of thermodynamic, regulatory, and reversibility constraints on the predictions from multiple genome-scale metabolic models.

3. Methods

3.1 Flux Balance Analysis (FBA)

Flux balance analysis (FBA) is a constraint-based simulation method used to define the limits on the metabolic capabilities of a microorganism [12-14]. In FBA, the interior of the cell is assumed to be in a quasi-steady-state, meaning that the net production/ consumption of each internal metabolite is zero. Based on this assumption, linear constraints are established on the flux through each reaction involved in the organism metabolism. Reaction fluxes are further constrained based on knowledge of the reversibility and directionality of the metabolic reactions, determined from thermodynamics [6-8]. A linear optimization is then performed

with these constraints, such that a given metabolic objective function (often cell growth [15]) is maximized subject to the mass balance constraints, the reversibility constraints, and the availability of nutrients in the media. Gene knockouts may also be simulated by blocking all flux through metabolic reactions that are associated with the knocked out genes. Media conditions are set by restricting the compounds that can be consumed from the environment by the model reactions.

3.2 Classification of Reactions Using Flux Variability Analysis

Flux variability analysis (FVA) is an FBA-based method for characterizing the multiple feasible states of genome-scale metabolic models and for classifying the model reactions according to their behavior during simulated growth [16]. The reaction classification is derived from the minimization and maximization of flux through each model reaction while constraining the biomass production in the model to a minimal growth rate. Reactions with a minimum and maximum flux of zero are classified as *blocked* in the simulated conditions; reactions with a negative maximum flux or positive minimum flux are classified as *essential* in the simulated conditions; and all other reactions are classified as *active*.

4. Results

4.1 Impact of Reversibility Constrains on Model Accuracy

In order to study the effect of reversibility constraints on the accuracy of genome-scale metabolic model predictions, two genome-scale metabolic models of *E. coli* K12 (*i*JR904 [4] and *i*AF1260 [11]) and one genome-scale metabolic model of *B. subtilis* 168 (*i*Bsu1103 [6]) were utilized to predict the outcome of gene essentiality and Biolog growth phenotype experiments. These models were selected for analysis because they represent two of the most-well-studied prokaryotic organisms, one gram positive and one gram negative. Also, genome-wide gene essentiality and Biolog phenotyping array data are readily available for both of these organisms. Essentiality data is available for *E. coli* K12 in three distinct media conditions: Luria-Bertani media, glucose minimal media, and glycine minimal media [17, 18]. Essentiality data is also available for *B. subtilis* 168 in one culture condition: Luria-Bertani media [19].

The metabolic models were utilized to perform gene knockouts *in silico*, while simulating all culture conditions where experimental data is available. Knockouts were performed while enforcing and relaxing (by making all reactions reversible) the reversibility constraints included in each model. Predictions were then compared with experimental data to assess accuracy with and without reversibility constraints (Table 1). In all three models, the accuracy of gene essentiality predictions declined significantly with the relaxation of reversibility constraints, while the accuracy in predicting nonessential genes increased only for the *i*JR904 and *i*AF1260 models. This relaxation of reversibility constraints consists of making all model reactions reversible.

Biolog phenotyping arrays [20] have also been constructed and utilized to study the ability of *E. coli* K12 and *B. subtilis* 168 to metabolize 324 and 242 distinct carbon, nitrogen, phosphate, and sulfate sources respectively. The *i*JR904, *i*AF1260, and *i*Bsu1103 models were used to replicate these Biolog growth conditions *in silico*, while enforcing and relaxing the reaction reversibility constraints; all predictions were then compared against the experimental Biolog data (Table 1). In these studies, the accuracy of all three models in predicting the metabolized Biolog nutrients improved with the relaxation of reversibility constraints, while accuracy in predicting un-metabolized nutrients declined. However, the improvement in the prediction of metabolized nutrients was much more substantial for the *i*JR904 and *i*AF1260 models than for the *i*Bsu1103 model.

Table 1. Accuracy of Model Predictions with and without Reversibility Constraints

	iJR904*		iAF1260*		iBsu1103*	
Reversibility constraints	ON	OFF	ON	OFF	ON	OFF
Biolog conditions with growth	77/194 (40%)	99/194 (51%)	114/194 (59%)	130/194 (67%)	138/169 (82%)	142/169 (84%)
Biolog conditions with no growth	106/130 (82%)	79/130 (61%)	98/130 (75%)	71/130 (55%)	68/73 (93%)	50/73 (68%)
Essential metabolic genes	340/518 (66%)	229/518 (44%)	392/615 (64%)	99/615 (16%)	192/215 (89%)	166/215 (77%)
Non-essential metabolic genes	2000/2137 (94%)	2057/2137 (96%)	3053/3165 (95%)	3155/3165 (100%)	873/888 (98%)	873/888 (98%)
Overall accuracy	82.1%	80.2%	89.1%	84.2%	94.5%	91.5%

*Gene K.O. simulation results represent the aggregate of 3 media conditions (Luria-Bertani media, glucose minimal media, and glycine minimal media [17, 18]).

4.2 Impact of Reversibility Constraints on Reaction Behavior

Another measure of model quality is the number of inactive reactions in the model. Many reactions are supposed to be inactive during growth on certain condi-

tions. For example, reactions involved in glycine metabolism should be inactive during growth on glucose minimal media. However, other reactions are inactive because they either exclusively lead to or are derived from a dead end in the metabolic network.

We utilized FVA to identify inactive reactions in the *i*JR904, *i*AF1260, and *i*Bsu1103 models during minimal simulated growth in *complete media*. In c*omplete media*, any transportable nutrient is allowed to be taken up by the cell, making it the least restrictive media condition possible. The advantage of performing FVA on complete media is that this enables as many reactions as possible to be active since no uptake pathways are blocked. Thus, reactions identified as inactive in complete media represent those reactions that will never carry flux because they exclusively lead to or are derived from a dead-end metabolite. In some cases, these dead-ends can be eliminated with the relaxation of reversibility constraints. To identify these dead-ends, we repeated the FVA reaction classification to identify reactions that are no longer inactive with reversibility constraints relaxed (Figure 1). In all three models, the number of inactive reactions declined substantially with the relaxation of the reversibility constraints.

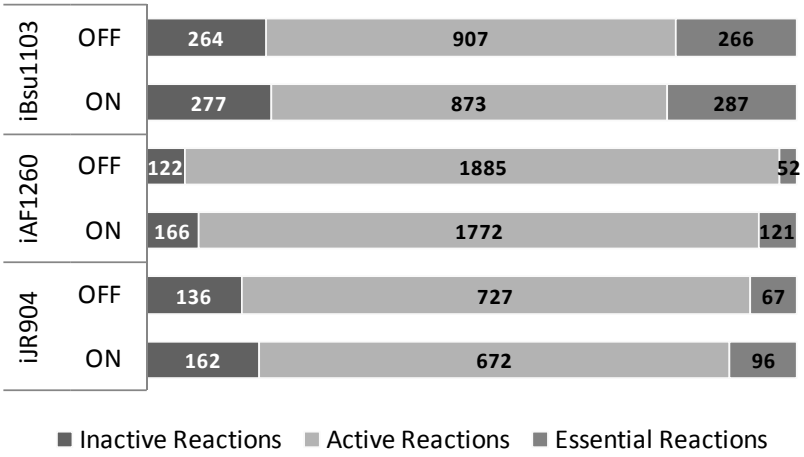

Fig. 1. Number of inactive, active and essential reactions with reversibility constraints turned "ON" and "OFF"

5. Discussion

The results of our analysis of the effect of reversibility constraints on the accuracy of model predictions demonstrated that complete relaxation of reversibility constraints always results in a substantial decline in accuracy. However, the results also reveal that many cases where no growth is predicted and growth is observed (false negative predictions) can be corrected with the relaxation of reversibility constraints alone. More rigorous optimization techniques are available [9, 10] for identifying exactly which reactions should be made reversible to correct these predictions; however, this simple study provides a bulk estimate of how effective such efforts will be and it identifies the exact conditions on which such efforts should be applied. This study also reveals the correctly predicted zero-growth conditions that are vulnerable to being broken by the adjustment of reversibility constraints. Both pieces of information can be used to substantially simplify procedures for optimizing reaction reversibility constraints in models to fit experimental data.

Another interesting result can be derived from contrasting the effect of the reversibility constraints on the *i*JR904 and *i*AF1260 models versus the *i*Bsu1103 model. While in all three models, the number of false negative predictions declined with the relaxation of reversibility constraints, the decline was much more substantial in the *E. coli* models compared with the *i*Bsu1103 model. Meanwhile, the rise in false positive predictions with the relaxation of reversibility constraints was comparable in all three models. This ratio of errors corrected over errors created with the relaxation of reversibility constraints can be used as a measure of the extent to which a genome-scale metabolic model has been optimized. Thus, the reversibility rules in the *i*Bsu1103 model, which was optimized during reconstruction using the *GrowMatch* method, show a greater extent of optimality compared with the reversibility rules in the *i*AF1260 and *i*JR904 models, which underwent manual optimization only.

6. Acknowledgements

This work was supported by the U.S. Dept. of Energy under Contract DE- AC02-06CH11357.

7. References

1. Feist AM and Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nat Biotechnol **26**(6):659-67
2. Feist AM, Herrgard MJ, Thiele I et al (2009) Reconstruction of Biochemical Networks in Microbial Organisms. Nature Reviews Microbiology **7**(2):129-143
3. Henry CS, Xia F, and Stevens RL (2009) Application of High-Performance Computing to the Reconstruction, Analysis, and Optimization of Genome-scale Metabolic Models. Journal of Physics: Conference Series **180**(012025)
4. Reed JL, Vo TD, Schilling CH et al (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biology **4**(9):1-12
5. Edwards JS and Palsson BO (2000) The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proceedings of the National Academy of Sciences of the United States of America **97**(10):5528-5533
6. Henry CS, Zinner J, Cohoon M et al (2009) iBsu1103: a new genome scale metabolic model of B. subtilis based on SEED annotations. Genome Biology **10**:R69
7. Kummel A, Panke S, and Heinemann M (2006) Systematic assignment of thermodynamic constraints in metabolic network models. BMC Bioinformatics **7**:512
8. Henry CS, Jankowski MD, Broadbelt LJ et al (2006) Genome-scale thermodynamic analysis of Escherichia coli metabolism. Biophysical Journal **90**(4):1453-61
9. Kumar VS and Maranas CD (2009) GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. PLoS Comput Biol **5**(3):e1000308
10. Kumar VS, Dasika MS, and Maranas CD (2007) Optimization based automated curation of metabolic reconstructions. BMC Bioinformatics **8**:212
11. Feist AM, Henry CS, Reed JL et al (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol **3**:121
12. Varma A and Palsson BO (1993) Metabolic capabilities of Escherichia-coli.1. Synthesis of biosynthetic precursors and cofactors. Journal of Theoretical Biology **165**(4):477-502
13. Varma A and Palsson BO (1993) Metabolic capabilities of Escherichia-coli. 2. Optimal-growth patterns. Journal of Theoretical Biology **165**(4):503-522
14. Papoutsakis ET and Meyer CL (1985) Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. Biotechnology and Bioengineering **27**(1):50-66
15. Edwards JS, Ibarra RU, and Palsson BO (2001) In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nature Biotechnology **19**(2):125-130
16. Chen T, Xie ZW, and Ouyang Q (2009) Expanded flux variability analysis on metabolic network of Escherichia coli. Chinese Science Bulletin **54**(15):2610-2619
17. Joyce AR, Reed JL, White A et al (2006) Experimental and computational assessment of conditionally essential genes in Escherichia coli. J Bacteriol **188**(23):8259-71
18. Baba T, Ara T, Hasegawa M et al (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2**:2006 0008
19. Kobayashi K, Ehrlich SD, Albertini A et al (2003) Essential Bacillus subtilis genes. Proc Natl Acad Sci U S A **100**(8):4678-83
20. Bochner BR (2009) Global phenotypic characterization of bacteria. Fems Microbiology Reviews **33**(1):191-205